Conditional independence and conditioned limit laws

Citation for published version:
Papastathopoulos, I 2016, 'Conditional independence and conditioned limit laws', *Statistics and Probability Letters*, vol. 112, pp. 1-4. https://doi.org/10.1016/j.spl.2015.12.028

Digital Object Identifier (DOI):
10.1016/j.spl.2015.12.028

Conditional independence and conditioned limit laws

Ioannis Papastathopoulos

School of Mathematics and Maxwell Institute, University of Edinburgh, Edinburgh, EH9 3FD

i.papastathopoulos@ed.ac.uk

Abstract

Conditioned limit laws constitute an important and well developed framework of extreme value theory that describe a broad range of extremal dependence forms including asymptotic independence. We explore the assumption of conditional independence of X_1 and X_2 given X_0 and study its implication in the limiting distribution of (X_1, X_2) conditionally on X_0 being large. We show that under random norming, conditional independence is always preserved in the conditioned limit law but might fail to do so when the normalisation does not include the precise value of the random variable in the conditioning event.

Key-words: Asymptotic independence; conditional independence; conditioned limit laws; random norming

AMS subject classifications: Primary: 60GXX, Secondary: 60G70

1 Introduction

Extreme value theory is a highly active area of research and its methods and applications are the epitome of risk modelling and statistical estimation of rare events. Extreme events that occur in a broad range of disciplines such as in environmental processes or in finance and insurance are typically multivariate in nature and usually have tremendous socio-economic impact. The recent technological advances have resulted in an ever-increasing amount of information available across the whole spectrum of applied sciences. As such, when modelling data in several dimensions, one is typically confronted by the *curse of dimensionality*. It is widely recognized that the construction of more efficient statistical models and techniques that overcome this problem is imperative. *Conditional independence* constitutes one of the most fundamental tools and concepts in this direction (Besag, 1974; Dawid, 1979; Whittaker, 1990; Lauritzen, 1996; Cox and Wermuth, 1996). On the other hand, the central concept of regular variation, its extensions and refinements, provide the recipe for the development of asymptotically justified extreme value models. The purpose of this short note is to illustrate some implications of conditional independence in conditioned limit laws (Heffernan and Tawn, 2004; Heffernan and Resnick, 2007), a key and well-developed framework that embodies a broad range of extremal dependence forms.

For ease of exposition, consider a random vector (X_0, X_1, X_2) in $\mathbb{R}^3$. The result developed in this paper extends to higher-dimensional settings for X_1 and X_2 in a straightforward manner so this is not restrictive. We assume that X_1 and X_2 are conditionally independent given X_0. Informally, this means that the conditional distribution of X_1 given (X_0, X_2) is equal to the conditional distribution of X_1 given X_0 alone. In other words, once X_0 is known any further information about X_2 is irrelevant to uncertainty about X_1 and therefore it readily follows that for any $x_1, x_2 \in \mathbb{R}$,

$$\mathbb{P}\left(X_1 < x_1, X_2 < x_2 \mid X_0\right) = \mathbb{P}\left(X_1 < x_1 \mid X_0\right) \mathbb{P}\left(X_2 < x_2 \mid X_0\right), \tag{1}$$

almost surely. We show that conditional independence may naturally be *preserved* in limiting laws of random vectors given an extreme component (Heffernan and Tawn, 2004; Heffernan and Resnick, 2007). Conditioned limit laws provide a rich description of extremes of random vectors that do not necessarily grow at the same rate and may exhibit *asymptotic independence* which means that the coefficient

$$\chi = \lim_{p \to 1-} \mathbb{P}\left[F_{X_1}(X_1) > p, F_{X_2}(X_2) > p \mid F_{X_0}(X_0) > p\right], \tag{2}$$

can be 0. Conditioned limit laws were systematically studied for first time by Heffernan and Tawn (2004) who examined the limiting conditional distribution of affinely transformed random vectors as the conditioning variable becomes large. Assuming identical margins with distribution function being asymptotically equivalent to the unit exponential distribution, i.e., $F_{X_0}(x) = F_{X_1}(x) = F_{X_2}(x) \sim 1 - \exp(-x)$, as $x \to \infty$, Heffernan and Tawn showed that for a broad range of dependence structures of (X_0, X_1, X_2), there exist scaling functions $\alpha_1, \alpha_2 : (0, \infty) \mapsto (0, \infty)$, location functions $\beta_1, \beta_2 : (0, \infty) \mapsto \mathbb{R}$ and a joint distribution G on $[-\infty, \infty) \times [-\infty, \infty)$ with non-degenerate marginals, such that as $t \to \infty$

$$\mathbb{P}\left[\frac{X_1 - \beta_1(X_0)}{\alpha_1(X_0)} < x_1, \frac{X_2 - \beta_2(X_0)}{\alpha_2(X_0)} < x_2 \;\middle|\; X_0 > t\right] \xrightarrow{\mathcal{D}} G(x_1, x_2), \tag{3}$$

on $[-\infty, \infty] \times [-\infty, \infty]$ where $\xrightarrow{\mathcal{D}}$ stands for weak convergence. Although the original formulation of Heffernan and Tawn relied on the existence of densities, formulation (3) is presented here in the compact form of Heffernan and Resnick (2007) who provided a formal mathematical examination of the more general conditioned limit formulation that there exists a joint distribution H on $[-\infty, \infty) \times [-\infty, \infty)$ with non-degenerate marginals such that as $t \to \infty$

$$\mathbb{P}\left[\frac{X_1 - \beta_1(t)}{\alpha_1(t)} < x_1, \frac{X_2 - \beta_2(t)}{\alpha_2(t)} < x_2 \;\middle|\; X_0 > t\right] \xrightarrow{\mathcal{D}} H(x_1, x_2), \tag{4}$$

on $[-\infty, \infty] \times [-\infty, \infty]$, subject to the sole assumption of X_0 belonging to the domain of attraction of an extreme value distribution. Expressions (3) and (4) can be rephrased more generally as special cases of joint probability convergence; here we use the conditional representation to highlight the connection with conditional independence.

Limit expressions (3) and (4) differ in the way X_1 and X_2 are normalised since in expression (3), the precise value of X_0 that occurs with $X_0 > t$ is used, whereas in expression (4) only partial information about the value of X_0 is exploited since only the level value t that X_0 exceeds is used. Following the terminology of Heffernan and Resnick (2007) we refer to limit expressions (3) and (4) as the conditional extreme value model with *random norming* and *deterministic norming*, respectively, and for the remaining part we assume without loss of generality that X_0 has a unit Pareto distribution, i.e., $\mathbb{P}(X_0 < x) = 1 - 1/x$, $x > 1$. The unit Pareto marginal scale for X_0 is primarily chosen for convenience but the result of this paper can be stated, with modified proofs (Kulik and Soulier (2015)), for X_0 being in the domain of attraction of an extreme value distribution. We also assume that the two limit expressions (3) and (4) simultaneously hold for the same pair of norming functions. In general, this may not always be true but a necessary and sufficient condition is to assume that (α_i, β_i), $i = 1, 2$, in expression (4) are extended regularly varying, i.e., there exist $\rho_1, \rho_2, \kappa_1, \kappa_2 \in \mathbb{R}$ such that as

$t \to \infty$,

$$\frac{\alpha_i(tx)}{\alpha_i(t)} \to x^{\rho_i} \qquad \text{and} \qquad \frac{\beta_i(tx) - \beta_i(t)}{\alpha_i(t)} \to \psi_i(x),$$

for $x > 0$, where

$$\psi_i(x) = \begin{cases} \kappa_i \left(x^{\rho_i} - 1\right)/\rho_i & \rho_i \neq 0 \\ \kappa_i \log x & \rho = 0, \end{cases} \tag{5}$$

for $i = 1, 2$ (Resnick and Zeber, 2014).

In Section 2 we state the main theorem of the paper which shows that the conditional independence property (1) is preserved in the conditional extreme value model with random norming, meaning that for any $x_1, x_2 \in \mathbb{R}$,

$$G(x_1, x_2) = G_1(x_1)\, G_2(x_2), \tag{6}$$

where $G_1(x_1) = \lim_{x_2 \to \infty} G(x_1, x_2)$ and $G_2(x_2) = \lim_{x_1 \to \infty} G(x_1, x_2)$, but might fail to do so in the conditional extreme value model with deterministic norming. In Section 3 we discuss practical consequences of conditional independence in conditioned limit laws. A proof is given in Section 4.

2 Main result

Let $\pi_1, \pi_2 : (0, \infty) \times \mathcal{B}(\bar{\mathbb{R}}) \mapsto [0, 1]$ be Markov kernels defined by

$$\pi_1(x, A) = \mathbb{P}(X_1 \in A \mid X_0 = x) \quad \text{and} \quad \pi_2(x, A) = \mathbb{P}(X_2 \in A \mid X_0 = x).$$

Let also $\sigma(X_0) = \{{X_0}^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$ be the sigma algebra generated by X_0.

Theorem 1. *Let (X_0, X_1, X_2) be a random vector from a probability space to $\mathbb{R}^3$ and suppose that given X_0, X_1 is conditionally independent of X_2, i.e.,*

$$\mathbb{E}\left\{f_1\left(X_1\right)\; f_2\left(X_2\right) \mid \sigma(X_0)\right\} = \mathbb{E}\left\{f_1\left(X_1\right) \mid \sigma(X_0)\right\} \mathbb{E}\left\{f_2\left(X_2\right) \mid \sigma(X_0)\right\}$$

almost surely, for any Borel measurable functions $f_1, f_2 : \mathbb{R} \mapsto \mathbb{R}$ for which $f_1 \circ X_1$ and $f_2 \circ X_2$ are Lebesgue integrable with respect to $\mathbb{P}$. Suppose there exist two pairs of scaling and location *functions* $(\alpha_1,\beta_1),(\alpha_2,\beta_2):(0,\infty)^2\mapsto(0,\infty)\times\mathbb{R}$ *that are extended regularly varying and families of non-degenerate probability distribution functions* $\{G_{v;1},v>0\}$ *and* $\{G_{v;2},v>0\}$ *on* $[-\infty,\infty]$*, such that as* $t\to\infty$

$$\pi_1(t,(-\infty,\alpha_1(t)x_1+\beta_1(t)]) \xrightarrow{\mathcal{D}} G_1(x_1) \quad \textit{and} \quad \pi_2(t,(-\infty,\alpha_2(t)x_2+\beta_2(t)]) \xrightarrow{\mathcal{D}} G_2(x_2), \tag{7}$$

on $[-\infty,\infty]$ *where* $G_1=G_{1;1}$ *and* $G_2=G_{1;2}$*. Then, as* $t\to\infty$ *and for* f_1 *and* f_2 *bounded*

$$\mathbb{E}\left\{f_1\left(\frac{X_1-\beta_1(X_0)}{\alpha_1(X_0)}\right)f_2\left(\frac{X_2-\beta_2(X_0)}{\alpha_2(X_0)}\right)\;\Big|\;X_0>t\right\}\to\mathbb{E}_{G_1}\{f_1(X_1)\}\mathbb{E}_{G_2}\{f_2(X_2)\}. \tag{8}$$

and

$$\mathbb{E}\left\{f_1\left(\frac{X_1-\beta_1(t)}{\alpha_1(t)}\right)f_2\left(\frac{X_2-\beta_2(t)}{\alpha_2(t)}\right)\;\Big|\;X_0>t\right\}\to\int_1^\infty\mathbb{E}_{G_{v;1}}\{f_1(X_1)\}\mathbb{E}_{G_{v;2}}\{f_2(X_2)\}v^{-2}\,dv \tag{9}$$

3 Discussion

Conditional independence shows that important simplifications can be achieved at a practical level. To elaborate on this, consider a regression setting where interest lies in estimating the conditional probability on the left hand side of equation (1). A statistician has virtually two options. The first is to estimate the joint conditional probability by means of either a parametric model or to proceed non-parametrically. The second approach consists of estimating in the same spirit the lower dimensional marginal conditional distributions and combine them multiplicatively to yield an estimate. Thus, in light of extra information regarding conditional independence, the second approach is more efficient since it relates to a lower dimensional problem.

Theorem 1 shows that the inclusion of the precise value of the random variable in the conditioning event adds enough detail to the normalisation to always allow the limit law to factorise under the presence of conditional independence. This would offer substantial simplifications in practical as well as theoretical extreme value analyses. On the other hand, it follows from Theorem 1 and Lemma 1 that

$$H(x_1,x_2)=\int_1^\infty G_1\left(\frac{x_1-\psi_1(v)}{v^{\rho_1}}\right)G_2\left(\frac{x_2-\psi_2(v)}{v^{\rho_2}}\right)v^{-2}\,\mathrm{d}v, \tag{10}$$

with ψ_1 and ψ_2 defined by equation (5). This implies that $H(x_1, x_2) = H_1(x_1)\, H_2(x_2)$ if and only if either $(\kappa_1, \rho_1) = (0, 0)$ or $(\kappa_2, \rho_2) = (0, 0)$. Generally, this condition is not always satisfied and this shows why random norming is essential for linking conditional independence to conditioned limit laws.

Last, we note that conditional independence might not always be an inherent feature of multivariate extreme value models (Papastathopoulos and Strokorb, 2016). Preservation of conditional independence in conditioned limit laws signifies a potential benefit with this framework and opens up a possible research direction.

4 Appendix: Proof of Theorem 1

The proof of Theorem 1 is based on Resnick and Zeber (2014) proposition that links kernel convergence to the conditional extreme value model with deterministic norming.

Lemma 1 (Proposition 4.1, Resnick and Zeber (2014)). *Let G be a non-degererate probability distribution function on $[-\infty, \infty)$ and $\pi : (0, \infty) \times \mathcal{B}(\bar{\mathbb{R}}) \mapsto [0, 1]$ be a transition function satisfying as $t \to \infty$*

$$\pi(t, [-\infty, \alpha(t)\, x + \beta(t)]) \xrightarrow{\mathcal{D}} G(x) \tag{11}$$

on $[-\infty, \infty]$, where $\alpha(\cdot) \in \mathbb{R}$ and $\beta(\cdot) > 0$ are location and scaling functions. There exists a family of non-degenerate probability distribution functions $\{G_v : 0 < v < \infty\}$ on $[-\infty, \infty)$ such that for $0 < v < \infty$ and as $t \to \infty$

$$K(tv, [-\infty, \alpha(t)\, x + \beta(t)]) \xrightarrow{\mathcal{D}} G_v(x)$$

on $[-\infty, \infty]$, if and only if α, β are extended regularly varying with parameters $\rho, \kappa \in \mathbb{R}$,

$$\frac{\alpha(tx)}{\alpha(t)} \to x^\rho \qquad \textit{and} \qquad \frac{\beta(tx) - \beta(t)}{\alpha(t)} \to \psi(x) = \begin{cases} \kappa\,(x^\rho - 1)\,/\rho & \rho \neq 0 \\ \kappa \log x & \rho = 0 \end{cases},$$

for $x > 0$. In this case, $G_1 = G$, and

$$\pi(tv_t, [-\infty, \alpha(t)\, x + \beta(t)]) \to G(\{x - \psi(v)\}/v^\rho),$$

at continuity points of G, whenever $v_t = v(t) \to v \in (0,\infty)$.

Proof of Theorem 1 The left hand side of expression (9) is equal to

$$\frac{1}{\mathbb{P}\left(X_0 > t\right)} \int_t^\infty \int_{\mathbb{R}\times\mathbb{R}} f_1\left(\frac{x_1-\beta_1(t)}{\alpha_1(t)}\right) f_2\left(\frac{x_2-\beta_2(t)}{\alpha_2(t)}\right) \mathbb{P}\left(X_0 \in dx_0, X_1 \in dx_1, X_2 \in dx_2\right)$$

$$= \int_1^\infty \left[\int_{\mathbb{R}} f_1\left(x_1\right) \left[\int_{\mathbb{R}} f_2\left(x_2\right) \pi_2\left(tv, \alpha_2(t)\, dx_2 + \beta_2(t)\right)\right] \pi_1\left(tv, \alpha_1(t)\, dx_1 + \beta_1(t)\right)\right] \frac{\mathbb{P}\left(X_0 \in t\, dv\right)}{\mathbb{P}\left(X_0 > t\right)}.$$

The term in the inner square brackets of the last expression is bounded and, by assumption (7) and Lemma 1, it converges to

$$\mathbb{E}_{G_{v;2}}\{f_2(X_2)\} = \int_{\mathbb{R}} f_2(x_2)\, G_{v;2}(dx_2),$$

where $\{G_{v;2} : 0 < v < \infty\}$ is a family of non-degenerate probability distributions on $[-\infty,\infty)$ such that $G_{1;2} = G_2$. Working similarly for the term in the outer square brackets, we get that the left hand side of expression (9) equals

$$\int_1^\infty \mathbb{E}_{G_{v;1}}\{f_1(X_1)\}\, \mathbb{E}_{G_{v;2}}\{f_2(X_2)\}\, v^{-2}\, dv \qquad \text{as } t \to \infty.$$

Expression (8) is obtained in a similar manner and this completes the proof.

Acknowledgments

Ioannis Papastathopoulos acknowledges funding from the SuSTaIn program - Engineering and Physical Sciences Research Council grant EP/D063485/1 - at the School of Mathematics, University of Bristol. The author would like to thank Sidney Resnick, Kirstin Strokorb and Jonathan Tawn for helpful discussions. Thanks to Cornell University, Universität Mannheim and Lancaster University for funding during visits in 2015.

References

Besag, J. (1974), 'Spatial interaction and the statistical analysis of lattice systems (with discussion)', *J. Roy. Statist. Soc., B* **36**, 192–236. 2

Cox, D. R. and Wermuth, N. (1996), *Multivariate Dependencies, Models Analysis and Interpretation*, Vol. 67 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London. 2

Dawid, A. P. (1979), 'Conditional independence in statistical theory', *J. Roy. Statist. Soc. Ser. B* **41**(1), 1–31. 2

Heffernan, J. E. and Resnick, S. I. (2007), 'Limit laws for random vectors with an extreme component', *Ann. Appl. Prob.* **17**, 537–571. 2, 3

Heffernan, J. E. and Tawn, J. A. (2004), 'A conditional approach for multivariate extreme values (with discussion)', *J. Roy. Statist. Soc., B* **66**, 1–34. 2

Kulik, R. and Soulier, P. (2015), 'Heavy tailed time series with extremal independence.', *Extremes* pp. 1–27. 3

Lauritzen, S. L. (1996), *Graphical models*, Vol. 17 of *Oxford Statistical Science Series*, The Clarendon Press, Oxford University Press, New York. Oxford Science Publications. 2

Papastathopoulos, I. and Strokorb, K. (2016), 'Conditional independence among max-stable laws', *Statistics and Probability Letters* **108**, 9–15. 6

Resnick, S. I. and Zeber, D. (2014), 'Transition kernels and the conditional extreme value model', *Extremes* **17**, 263–287. 4, 6

Whittaker, J. (1990), *Graphical Models in Applied Multivariate Statistics*, Wiley. 2